

УДК: 681.58

к.геогр.н., с.н.с Путренко В.В.,
putrenko@wdc.org.ua, 0000-0002-0239-9241,
Назаренко С.Ю., nazarenko@wdc.org.ua 0000-0003-3367-5875,
Світовий центр даних з геоінформатики та сталого розвитку
НТУУ «КПІ ім. Ігоря Сікорського»,
к.геогр.н. Пашинська Н.М., n_pashynska@ukr.net, 0000-0002-0133-688X,
Київський національний університет імені Тараса Шевченка

МОДЕЛЮВАННЯ ДАНИХ З ПРОСТОРОВО-ЧАСОВИМ РОЗПОДІЛОМ ДЛЯ ПОТРЕБ ПРИЙНЯТТЯ РІШЕНЬ

Запропоновано наукові та практичні засади аналізу даних за допомогою методології побудови просторово-часового кубу, як один із видів інтелектуального аналізу даних з просторово-часовим розподілом. Використання даного методу на прикладі аналізу інформації від абонентів однієї з великих телекомунікаційних мереж мобільного оператора дозволяє проводити статистичний аналіз та виявляти статистично значущі просторово-часові кластери у даних, що може бути використано під час структурування даних з метою забезпечення безпеки та швидкого реагування на небезпечні ситуації.

Ключові слова: просторово-часовий куб, ГІС-аналіз, візуалізація даних, прийняття рішень.

Постановка проблеми. Сучасний світ надзвичайно багатий найрізноманітнішою інформацією, значні масиви якої людина збирає, зберігає, аналізує, та на її основі намагається робити прогнози та передбачення. Особливо це стосується тої інформації котра на пряму та опосередковано впливає на безпеку людини. Не менш важлива інформація для органів управління, правильний аналіз якої дає змоги зробити крок у вірному (прогресивному) напрямку. Але оперування великими масивами даних, що розосереджені не лише по значній території, але і у великому часовому інтервалі зазвичай ускладнюють прийняття правильного рішення, або ж значно збільшують час на його прийняття. Не менш важливим є аналіз попередніх даних, та їх розподіл у просторі та часі, за для прогнозування та передбачення майбутніх надзвичайних ситуацій чи катастроф. Окрім цього швидке реагування на надзвичайну ситуацію значно збільшує шанси на зменшення кількості жертв або іншого негативного наслідку, починаючи від конкретної людини до масштабів держави. Також просторово-часові дані, що дозволяють визначити переміщення різноманітних потоків (товарних, фінансових,

транспортних, людських) та сформувані їх закономірності мають важливе значення для просторового планування.

Огляд попередніх досліджень. Останнім часом використання великих масивів даних, що розосереджені не лише по значній території, але і у великому часовому інтервалі, потребує відповідних технологій для прийняття правильного рішення. Одним з варіантів вирішення даного питання є дослідження, аналіз та прогнозування події за допомогою побудови просторово-часового кубу. Вперше використання просторово-часового кубу запропонував Т. Хейгерstrand на початку 70-х років [1], можливості якого він описав у своїй праці «What about people in regional science?». Але до активного розвитку геоінформаційних систем (ГІС) його використання було обмежене. Лише в 2000 – х роках з'являються роботи з використання просторово-часового кубу в ГІС. У цих роботах представлялися нові можливості використання просторово-часового кубу за допомогою ГІС, в тому числі включаючи дослідження землетрусів [2-4].

Наступними кроками у використанні методу просторово-часового кубу було його застосування в інтелектуальному аналізі даних різноманітної природи: аналіз злочинів, вивчення інфраструктури, аналіз поведінки тварин, візуалізації руху людини, дослідження залежностей зміни погодних умов в часі [5-7]. В Україні лише деякі роботи лише частково зачіпають тему використання просторово-часового кубу, але не розкривають її повністю [8-10]. Таким чином, моделювання великих масивів багатовимірних даних, їх аналіз та використання для потреб прийняття рішень у різних сферах є важливим завданням, що потребує вирішення.

Мета та завдання. Метою роботи є аналіз просторово-часових закономірностей у розподілі подій у мережі Vodafone на основі використання методології побудови просторово-часового кубу.

Завданнями є:

- дослідження методології використання просторово-часового кубу для інтелектуального аналізу просторово-часових даних;
- вивчення особливостей застосування методу побудови просторово-часового кубу для аналізу просторово-часових рядів даних, що генеруються користувачами телекомунікаційної мережі Vodafone;
- застосування побудови просторово-часового кубу для аналізу розподілу просторових та часових паттернів даних мобільного зв'язку з метою екстреного реагування на надзвичайні події природного та соціального характеру.

Виклад основних результатів.

Просторово-часовий куб - це техніка 3D-візуалізації, призначена для одночасного представлення просторових та часових характеристик руху. Згідно з цим, точки траєкторій відображаються у тривимірному просторі, де вертикальна вісь зазвичай виражає час [10].

На початку 70-х років Т. Хейгерstrand [1] розробив графічний погляд на час як додатковий просторовий вимір. Він запропонував тривимірну діаграму, так званий просторово-часовий куб, щоб показати життєві історії людей і те, як люди взаємодіють у просторі та часі. Площина куба являє собою двовимірний географічний простір, а висота куба - часовий вимір. У той час, коли була представлена концепція, варіанти створення такої графіки обмежувались ручними методами, що слугувало очевидним обмеженням підходу, оскільки створення кожної нової діаграми виявилось трудомістким заняттям. Сьогодні ж, коли сучасні комп'ютерні технології мають набагато кращі можливості для візуалізації даних, графічний погляд Т. Хейгерstrandа отримує друге життя та активно застосовується у ГІС.

Використання просторово-часового кубу потребує просторові та часові дані для аналізу певних подій. Прикладами таких подій є землетруси, дорожньо-транспортні пригоди, випадки захворювань або спостереження рідкісних тварин [4].

Т. Хейгерstrand пропонував застосовувати просторово-часовий куб до даних про рух об'єктів, тобто про зміни просторових місць з прив'язкою до часу. У даній роботі автори пропонують застосувати концепцію Т. Хейгерstrandа до іншого типу даних, а саме до аналізу мережевих подій. Окрім цього використання просторово-часового кубу дає змогу відповісти на 3 питання Пьюке, поставлені до просторово-часових даних [11]:

- коли + де → що: опис об'єктів або набір об'єктів, які присутні в певному місці або наборі місцеположень за певний час або проміжок часу;
- коли + що → де: опис місця розташування або набір місць розташування, зайнятих певним об'єктом або набором об'єктів у певний час або проміжку часу;
- де + що → коли: опис конкретного часу або проміжку часу, коли певний об'єкт або набір об'єктів займали певне місце або набір місць розташування.

Для проведення дослідження використовуються дані надані компанією Vodafone, котрі мають просторову та часову прив'язку, а також певну атрибутивну інформацію. Опрацьована база даних нараховує 150 тисяч дзвінків, повідомлень та виходів у інтернет, з найрізноманітніших пристроїв та

від різних абонентів. Усі «події» зосереджені практично у західних областях України.

За для досягнення поставлених цілей, автори використовують набір інструментів Поглибленого аналізу просторово-часових закономірностей у програмному забезпеченні ArcMap 10.5. Даний набір інструментів містить статистичні інструменти для аналізу розподілу даних і виявлення закономірностей в контексті простору-часу. У набір входять інструменти: Створити куб Простір-Час та Аналіз виникнення гарячих точок.

Інструмент Створити куб Простір-Час на основі набору точкових даних будує тривимірну кубічну структуру (в форматі netCDF) для подальшого аналізу. Інструмент Аналіз виникнення гарячих точок використовує цей куб в якості вхідних даних і на основі нього виявляє статистично значущі тренди гарячих і холодних точок у часі. Інструмент Аналіз виникнення гарячих точок можна використовувати при обробці даних про правопорушення або спалахи епідемій для виявлення нових, зростаючих, спорадичних або постійних гарячих точок, досліджуючи серії тимчасових інтервалів. [12]

Інструмент Створити куб Простір-Час використовує точкові об'єкти з часовим атрибутом і структурує їх в куб формату netCDF, агрегуючи точки в біни простору-часу (Рис. 1). Для значень бінів виконується підрахунок кількості точок і визначається наявність трендів в часі в кожному місцезнаходження з використанням статистики Манна-Кенделла. [13]

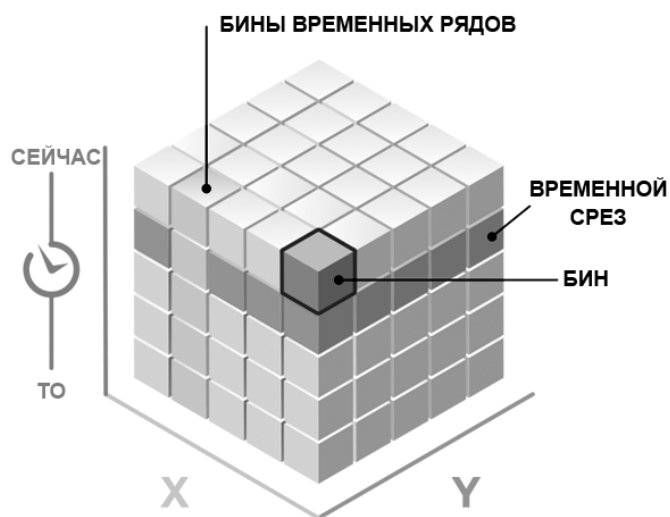


Рис. 1. Відображення бінів у просторово-часовому кубі (джерело: Arcmap Help [12])

Структура куба містить рядки, стовпці і часові кроки. Перемноживши кількість рядків на число стовпців і на число часових кроків, отримуємо сумарне число бінів у кубі. Рядки і стовпці визначають просторовий екстент куба, а часові кроки визначають часовий екстент. Місця розташування з даними

(біни в просторовому вимірі) - це області, розташовані в межах екстента, де протягом усього часового інтервалу зафіксовано хоча б одна точкова подія. [13]

Для більшості інструментів аналізу тільки місце розташування з даними хоча б для одного часового кроку буде включено в аналіз, при цьому аналіз буде проводитися для всіх часових кроків. Нульову кількість буде присвоєно біну, що не містить жодної точки, якщо у відповідному місцезнаходженні хоча б один часовий крок буде містити хоча б одну точку. Інформація про відсоток бінів з нульовою кількістю, пов'язаних з місцями розташування, для яких є дані хоча б для одного часового кроку, буде відображена в повідомленні, як величина розрідженості. [13]

Вхідні об'єкти повинні бути виключно точковими об'єктами, які представляють дані подій, наприклад, точки злочинів або пожеж, точок спалаху захворювання або дорожніх пригод, у нашому випадку подія, що пов'язана із вхідними чи вихідними дзвінками та повідомленнями, а також виходами користувачів у інтернет. У кожній точки повинна бути прив'язка до дати. Поле зі значенням часу для події повинно мати тип «Date». Ще однією умовою даного інструменту є наявність, як мінімум, 60 точок з різними значеннями часу. Даний інструмент припинить роботу, якщо в ході роботи буде створено куб з числом елементів, що перевищує два мільярди елементів (бінів). Для точного вимірювання відстаней цьому інструменту потрібні дані у прямокутній системі координат.

Разом із створеним просторово-часовим кубом інструмент надає вихідне повідомлення, у кінці якого знаходиться інформація про загальний тренд в даних. Тренд обчислюється на основі аналізу часових рядів в просторі. Основне питання, на яке відповідає аналіз - кількість подій, які представлені вхідними точками, збільшується або зменшується з плином часу? Щоб отримати відповідь на це запитання, значення кількості точок за місцем розташування групуються в тимчасові ряди і аналізуються з використанням статистики Манна-Кендалла [14].

Непараметричний тест Манна-Кендалла зазвичай використовується для виявлення тенденцій в серії даних. Нульова гіпотеза H_0 полягає в тому, що вибірка даних хронологічно впорядкована, незалежна і ідентично розподілена. Альтернативна гіпотеза H_A полягає в тому, що дані відповідають однозначному розподілу. Статистика тесту Манна-Кендалла розраховується за:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(X_j - X_k)$$

Де

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

Горизонтальні розміри біна за замовчуванням розраховуються наступним чином [13]:

- Обчислюється довжина більш довгої сторони екстента вхідних об'єктів (максимальний екстент).
- Далі розмір біна розраховується або як значення максимального екстента, розділене на 100, або за алгоритмом, заснованому на просторовому розподілі вхідних об'єктів (береться більше значення).

Інтервал часового кроку за замовчуванням розраховується з використанням двох різних алгоритмів, що визначають оптимальну кількість і ширину інтервалів часового кроку. Мінімальна кількість більше 10, отримане в результаті розрахунків з використанням двох алгоритмів, використовується як значення за замовчуванням для інтервалу часового кроку. Якщо в результаті обох алгоритмів отримано число менше 10, значення за замовчуванням для інтервалу часового кроку встановлюється на 10.

Отриманий набір даних Vodafone знаходиться у часовому відрізку з 1 червня 2017 року по 31 серпня 2017 року. За для зручності аналізу було використано 5-ти денний часовий крок. У наслідок чого інструментом було побудовано куб висотою у 19 бінів (Рис. 2).



Рис. 2. 3D візуалізація просторово-часового кубу, західних регіонів України

Інструмент Аналіз виникнення гарячих точок ідентифікує тренди в кластері щільності точок (підрахунків) або поля сум в просторово-часовому кубі, створеному за допомогою інструменту Створити куб простір-час. До категорій холодних і гарячих точок відносяться наступні характеристики [14]:

нова, послідовна, зростаюча, постійна, спадна, спорадична, що коливається і історична.

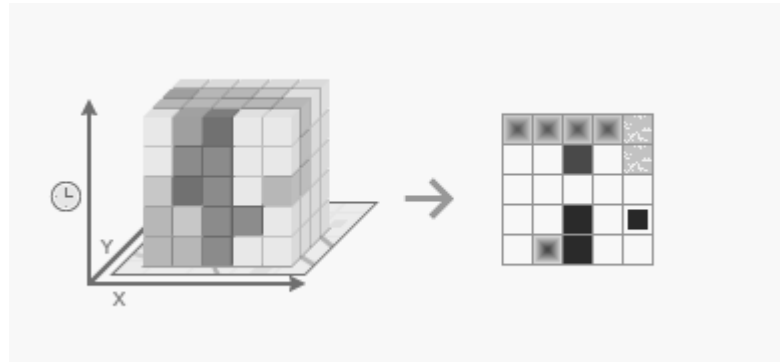


Рис. 3. Трансформація просторово-часового кубу для аналізу через гарячі точки (джерело: Arcmap Help [13])

Для цього інструменту необхідний файл netCDF, створений інструментом Створити куб Простір-час.

Кожен бін у просторово-часовому кубі містить значення LOCATION_ID, time_step_ID, COUNT і, можливо, поля підсумовування. Набір бінів, пов'язаний з одним і тим же місцем розташування, має однаковий ідентифікатор місця розташування і являє собою тимчасовий ряд. Набір бінів, пов'язаний з одним і тим же тимчасовим інтервалом, має однаковий ідентифікатор часового кроку і являє собою часовий зріз. Значення кількість в кожному бині представляє число точок, які присутні на певній території та певному часовому інтервалі.

Даний інструмент аналізує варіабельність у Вхідному кубі простору-часу netCDF, використовуючи просторово-часову інтерпретацію статистики Getis-Ord G_i^{**} , при цьому обчислюється значення кожного бина, яке зіставляється зі значеннями в сусідніх бинах.

Вихідні об'єкти додаються до Таблиці змісту і являють узагальнений результат просторово-часового аналізу для всіх проаналізованих місцеположень. Крім створення класу Вихідних об'єктів, зведені результати аналізу записуються в вікно Результати. Основним результатом використання інструменту Аналіз виникнення гарячих точок є створення карти, за допомогою якої надалі можливе дешифрування та знаходження гарячих точок (Рис. 4).

Серед інструментів Поглибленого аналізу просторово-часових закономірностей є 3-й інструмент: Аналіз локальних викидів, який так само як і попередній дозволяє дешифрувати важливі показники у Просторово-часовому кубі (Рис.5). Для ідентифікації кластерів та викидів у просторово-часовому кубі, даний інструмент використовує просторово-часову інтерпретацію статистики Anselin Локальний індекс Морана I, при цьому значення кожного бина зіставляється зі значеннями в сусідніх бинах.

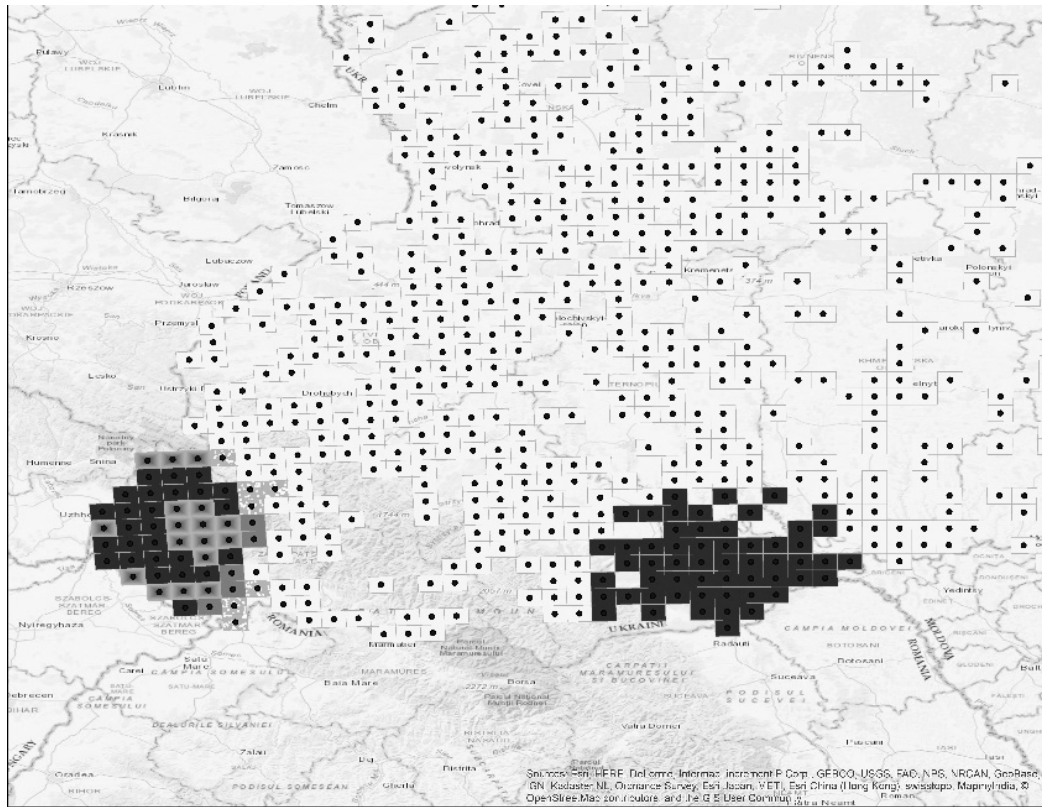


Рис. 4. Карта виникнення гарячих точок, створена в результаті аналізу просторово-часового кубу

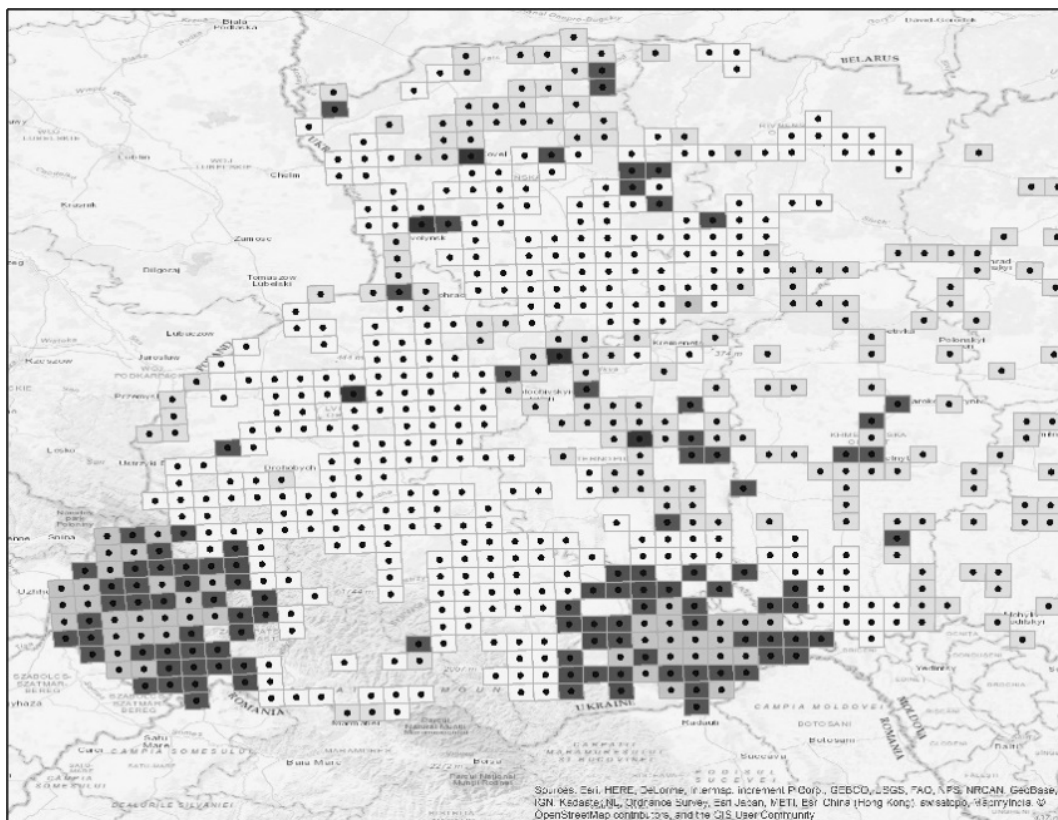


Рис. 5. Карта локальних викидів

Щоб ідентифікувати біни, які будуть включені в кожен околицю аналізу, інструмент спочатку виявляє біни, які потрапляють в задані параметри аналізу. Потім для кожного біна визначаються біни, розташовані в тому самому розташуванні, але в межах N попередніх тимчасових кроків, де N - Тимчасової крок околиці, зазначений у вхідних параметрах.

Біни з високими значеннями локальних викидів містять аномальні зміни у поведінці користувачів, які можуть мати різну природу як позитивного, так і негативного характеру. Разом з використанням класифікаторів та соціальних каналів розповсюдження новин, вони можуть бути ідентифіковані та передані до відповідних державних органів та служб.

Висновки.

Інструментарій побудови просторового-часового кубу надає зручний візуальний інтерфейс інтелектуального аналізу великих обсягів даних. Використання просторового-часового кубу можливо практично у будь-яких сферах, де необхідно аналізувати поведінку об'єктів та події, що відбуваються зі зміною місцеположення у просторі та часі.

Приклад використання просторово-часового аналізу даних для даних подій у мобільних мережах на прикладі мережі Vodafone надає змогу більш ефективно використовувати отримані дані, перш за все у безпекових цілях, що буде корисно урядовим організаціям для швидкого виявлення або попередження небезпечних ситуацій (таких як тероризм, надзвичайні явища, катастрофи чи інше). Надалі за допомогою просторово-часового кубу на основі даних мобільних операторів можна аналізувати статистичні викиди в активності абонентів у дзвінках або підключенню людей до інтернет з прив'язкою до певної території, що дасть змогу ідентифікувати певні аномалії та відповідно реагувати на ситуацію.

Список використаної літератури

1. Hägerstrand T. What about people in regional science? / T. Hägerstrand // Papers, Regional Science Association. – 24. - 1970. – pp. 7-21.
2. Kraak M. The Space Time Cube Revisited from a Geovisualization Perspective / M. Kraak // Proc. 21st Int'l Cartographic Conf. – 2003. - pp. 1988-1996.
3. Andrienko N. Exploratory SpatioTemporal Visualization: An Analytical Review / N. Andrienko, G. Andrienko, P. Gatalsky // J. Visual Languages and Computing. - vol. 14. - no. 6, 2003. - pp. 503-541.
4. Gatalsky P. Interaction Analysis of Event Data Using Space-Time Cube / P. Gatalsky, N. Andrienko, G. Andrienko // Proc. IEEE Eighth Int'l Conf. Information Visualization (IV '04). – 2004. - pp. 145-152.

5. Cheng T. Detecting emerging space-time crime patterns by prospective STSS / T. Cheng, M. Adepeju // Proc of the 12th International Conference on GeoComputation. Available: <http://www.geocomputation.org/2013/papers/77.pdf>. Accessed 2014 Jan 31
6. Gonçalves T. Visualizing Human Trajectories: Comparing Space-Time Cubes and Static Maps / T. Gonçalves, A.-P. Afonso, B. Martins // Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014 - Sand, Sea and Sky - 2014. – pp. 102-107
7. Yusof N. Mining frequent spatio-temporal patterns in wind speed and direction / N. Yusof, R. Zurita-Milla, M. J. Kraak // Connecting a digital Europe through location and space (Lecture notes in geoinformation and geography). – 2014. – pp. 143-161
8. Хурцилава К. В. К вопросу о пространственно-временных ГИС и некоторых их применениях / Хурцилава К. В., Білецький Б.О., Беспалов В.П. // Системи підтримки прийняття рішень. Теорія і практика. – Київ: ІПММС НАНУ. – 2013. – С. 52-55.
9. Путренко В.В. Застосування інструментарію геоінформаційного моделювання для інтелектуального аналізу даних пожежної небезпеки / В.В. Путренко, Н.М. Пашинська // Вісник Нац. ун-ту "ХПІ" : зб. наук. пр. Сер. : Нові рішення в сучасних технологіях. – Харків : НТУ "ХПІ", 2017. – № 7 (1229). – С. 156-163.
10. Путренко В. Використання 3D Space-time cube для моделювання просторово-часових даних / В. Путренко, С. Назаренко, Н. Пашинська // Збірник матеріалів конференції «ГІС-ФОРУМ-2018» — X. : ХНУ імені В. Н. Каразіна, 2018. — С. 32-34.
11. D.J. Peuquet. It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems / D.J. Peuquet // Annals of the Association of American Geographers. - 84 (3). -1994. – P. 441-461.
12. ESRI, Help Arcmap / An overview of the Space Time Pattern Mining toolbox – Available: <http://desktop.arcgis.com/ru/arcmap/10.3/tools/space-time-pattern-mining-toolbox/an-overview-of-the-space-time-pattern-mining-toolbox.htm>
13. ESRI, Help Arcmap / Create Space Time Cube – Available: <http://desktop.arcgis.com/ru/arcmap/10.3/tools/space-time-pattern-mining-toolbox/create-space-time-cube.htm>
14. ESRI, Help Arcmap / Create Space Time Cube – Available: <http://desktop.arcgis.com/en/arcmap/10.3/tools/space-time-pattern-mining-toolbox/emerginghotspots.htm#>

к.геогр.н., с.н.с. Путренко В.В., Назаренко С.Ю.,
Мировой центр данных по геоинформатике и устойчивому развитию
НТУУ «КПИ им. Игоря Сикорского»,
к.геогр.н. Пашинская Н.Н., Киевский национальный
университет имени Тараса Шевченко

МОДЕЛИРОВАНИЕ ДАННЫХ С ПРОСТРАНСТВЕННО- ВРЕМЕННЫМ РАСПРЕДЕЛЕНИЕМ ДЛЯ ПРИНЯТИЯ РЕШЕНИЙ

Предложены научные и практические основы анализа данных с помощью методологии построения пространственно-временного куба, как один из видов интеллектуального анализа данных с пространственно-временным распределением. Использование данного метода на примере анализа информации от абонентов одной из крупных телекоммуникационных сетей мобильного оператора позволяет проводить статистический анализ и выявлять статистически значимые пространственно-временные кластеры в данных, что может быть использовано при структурировании данных с целью обеспечения безопасности и быстрого реагирования на опасные ситуации.

Ключевые слова: пространственно-временной куб, ГИС-анализ, визуализация данных, принятия решений.

Ph.D., V. Putrenko, S. Nazarenko,

¹World Data Center for Geoinformatics and Sustainable Development
NTUU “Igor Sikorsky Kyiv Polytechnic Institute”,
Ph.D. N. Pashynska Taras Shevchenko National University of Kyiv

MODELING SPACE-TIME DATA FOR THE NEED FOR DECISION- MAKING

The scientific and practical principles of data analysis are proposed with the help of methodology of space-time cube construction as one of the types of intellectual analysis of data with spatial-temporal distribution. The use of this method on the example of information analysis from subscribers of one of the major telecommunication networks of a mobile operator allows to carry out a statistical analysis and to detect statistically significant spatio-temporal clusters in data that can be used during data structuring in order to provide security and react quickly to hazardous situations.

Keywords: space-time cube, GIS-analysis, data visualization, decision making.